DOCUMENT RESUME

ED 033 417 CG 004 585

AUTHOR Campbell, David P.
TITLE Some Desirable Characteristics of Interest Inventories.
INSTITUTION American Personnel and Guidance Association, Washington, D.C.
Pub Date Apr 69
Note 17p.; Paper presented at American Personnel and Guidance Association Convention, Las Vegas, Nevada, March 30-April 3, 1969

EDRS Price EDRS Price MF-$0.25 HC-$0.95
Descriptors Evaluation, Interests, *Interest Tests, *Test Construction, Test Interpretation, Test Reliability, Test Results, Test Selection, Test Validity, Theories, Vocational Interests

Abstract

A list of 11 important characteristics of interest inventories is presented. Comparisons between interest inventories are difficult to make, and those responsible for making choices between inventories must be well-informed about the entire system. Therefore, the points are discussed in the order that a system is developed--from the item pool, through the scale building, to the supporting data, through the theory to the application and to the commercialization of the product. Items should be drawn from the relevant domain should be in good taste and the item format should be simple and direct but allow the respondent some latitude in his answer. The interest inventory scales should be valid, stable over time and numerous and specific enough to allow easy interpolation yet few and broad enough to permit parsimonious generalizations. There should be an extensive body of published information on the inventory and the interest inventory should be tied into a theory. The results should be reported in an easily understandable manner and provide the individual with sufficient information about himself. Finally the interest inventory must be commercially viable, and the publishers should supply a wide array of ancillary services. (Author/RM)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED033417

Some Desirable Characteristics of Interest Inventories*

David P. Campbell
University of Minnesota

Comparisons between interest inventories are difficult to make. Those most informed about each inventory, that is, the authors, are the most partial and thus their conclusions must be treated gingerly; the data available for comparisons are hardly ever constant across inventories; and characteristics of inventories differ markedly in nature and are difficult to equate, e.g., how does one integrate information such as:

"Inventory A is inexpensive to score."

"Inventory B has 30 day test-retest reliabilities of only .65"

"Inventory C is fun to take."

"Inventory D has an overly-flamboyant advertising program."

The only solution is for those responsible for making choices between inventories to be well-informed about the entire system, and to make their decisions on characteristics that are important in their settings. For that purpose, the following list of important characteristics has been developed.

The points are listed here, not in order of importance, but rather in the order that a test system is developed--from the item pool, through the scale building, to the supporting data, through the theory, to the application, and finally to the commercialization of the product, for when one talks about the desirable characteristics of an interest inventory, one must consider all of these aspects.

---

*Paper presented at the meetings of the American Personnel and Guidance Association, Las Vegas, April, 1969.

CG004525

ITEM CONTENT

I. The inventory items should be drawn from the relevant domain, in this case, vocational activities, and within this domain provide a great deal of diversity.

The item content is extremely important for this determines the basic data that one has to work with. Although a few psychologists have offered the curious argument that content is unimportant--that factors such as response set, or response deviancy, or social desirability, or acquiescence are what determine responses--anyone who studies item responses of criterion groups even briefly quickly becomes convinced that item content is of overwhelming importance.

The items should cover broadly the domain one is trying to deal with; the difficulty here is that we don't know precisely what domain we are trying to cover, and we have even less knowledge about how various items map into this domain. For example, items about mechanical activities should be included for mechanical interests have some clustering integrity in that they hold together statistically as a focus of attention for people's preferences. Because tools are within that domain, some items should allow the respondent to express his feelings toward working with tools. Yet, even this straightforward area can be complicated, as a clever example from Holland's book on vocational choice (1966) demonstrates; he has cited two items dealing with tools that reflect two vastly different preferences:

The first item is: "I like to use tools to build things with."

The second is: "I like to use tools to hit people with."

(One can push the example into further complications with the item: "I like to use tools to build things to hit people with.")

Despite such problems, most of the well-known inventories have a decently broad coverage, derived--for the most part--from the common sense of their authors. While they have gaps--for example, the men's Strong Vocational Interest Blank has no items about homemaking, and some men would report preferences for activities such as "Decorating a room," "Preparing exotic food for a large dinner party," or "Helping a small child learn about nature" if given the opportunity--still the coverage in most inventories is adequate, if not as extensive as it might be.

Two reasons for having diverse item content involve the unpredictability of the future. The first is to fulfill future demands for new areas to be represented; when Strong made up his original items in the 1920's, he could hardly imagine that they would be used, 40 years later, to tap the interests of computer programmers and astronauts. The second "unpredictable" reason for diversity is that some items will inevitably become obsolete and will have to be discarded; Strong, in the 1920's, could not have predicted that the item "Attend vaudeville shows" would sound painfully dated in the 1960's, while other recreational items such as "Attend symphony concerts" or "Go camping" would survive.

II. <u>The items should be in good taste and offend as few people as possible</u>.

For years, the SVIB has had items asking the respondents to report their preferences for various types of people, such as:

"Deaf Mutes"

"Men with gold teeth"

"Negroes."

I think those items are in bad taste, and they have been removed. Further, in the recent revisions, the other items about various kinds of people have been revised, so that they are more closely related to occupational activities. Some of the new items are:

"Highway construction workers"

"High school students"

"Jet pilots"

"Girls who enter beauty contests."

Perhaps items about different types of people shouldn't be included at all, as they have been the most controversial portion of the inventory, yet responses to such items are clearly related to vocational preferences. For example, in response to the item, "Religious people," three times as many music teachers and five times as many Catholic sisters respond "Like" as do women psychologists. Differences of that magnitude are worth paying attention to, which is why that section of the SVIB was modified instead of discarded.

III. <u>The item format should be as simple and direct as possible, yet allow the respondent some latitude in his answer</u>.

The simplest format is probably the true-false question such as

the MMPI uses, but that permits only two choices and most people prefer more alternatives. The SVIB has three categories of response: Like, Indifferent, or Dislike, which is slightly better, but some people still complain that while they like some activities, they really like others and want some way to indicate this difference in attraction. I believe a five-choice item would be even better; E. K. Strong, I am told by one of his students, also believed this, though, for now, the SVIB is locked into three categories.

The forced-choice format, such as the triads on the Minnesota Vocational Interest Inventory where one is forced to choose between three items, or the section on the SVIB where the individual is confronted with ten activities and forced to choose the three most liked, is an undesirable form. Forced choice items are irritating to respondents, they are difficult to deal with psychometrically, and there are no data to support their use over other simpler forms.

I heard a testing psychologist argue in a symposium recently, seriously and with a straight face, that we should use forced choice items, and ipsitive scales, because life itself is a series of forced choices. Reasoning by analogy is usually dubious, and this was another example. He might as well argue that because life has a large measure of grief, we should add more sorrowful items to our inventories, or that because most of us spend more time in bed than any other place, the majority of items on an interest inventory should be concerned with bedroom activities.

The question as to which item format is best for which purpose is an empirical one, easily accessible to study, and test authors should do some research on their item forms, or at least read the existing literature, before offering simple speculation about which formats are best. There have been several studies comparing forced-choice with free-response formats (e.g., Berry (1955;Zuckerman, 1952); none has supported the use of either form over the other. At this point in history, anyone recommending forced-choice items over others has a considerable responsibility to present data, not analogies.

SCALE CONSTRUCTION

IV. Interest inventory scales should be stable over time.

Because these instruments are used to help individuals make relatively long term decisions, there must be some guarantee that their results are not ephemeral.

The only way to know which inventory is superior here is to have similar data for all on comparable samples. Such information is not available now for many inventories, and probably never will be, for the accumulation of such data requires a terrific investment of time, money, and psychic energy by the investigator. E. K. Strong was unique in his passion for longitudinal studies and it is unlikely that any other inventory will ever have comparable stability data collected over 40 years. This doesn't mean that other inventories are not equally stable; what it means is that we cannot tell.

V. Interest inventory scales should be valid.

There are many definitions of validity; the one that I prefer here is the ability of the inventory to separate disparate groups from each

other by magnitudes that have practical meaning--at least one or two standard deviations; otherwise the overlap between distributions is so large as to make the results meaningless. Definitions of validity that depend on statistical significance are especially irrelevant here, as the literature abounds with misleading reports of statistically significent, but trivial, results. Practical significance is what is important, and that can best by determined by looking at the magnitudes of group separation. The SVIB scales, on the average, separate occupations by about two standard deviations. Between extreme groups such as artists and Army officers, or policemen and psychologists, the separation approaches four standard deviations. Because so few test manuals present such data, comparisons are difficult.

VI. Interest inventory scales should be numerous and specific enough to allow easy interpretation, yet few and broad enough to permit parsimonious generalizations.

These conflicting goals are well represented by the extremes of the 156 scales for the Kuder OIS versus the six scales for Holland's VPI that represent the six types of people that John Holland's world is populated with.

On the SVIB, we are trying to have it both ways with two kinds of scales, the Occupational Scales and the new Basic Interest Scales. These represent the two kinds of systems, "open" and "closed," originally discussed by Clark (1961). The "open" system can continually have new scales added as new criterion groups are tested, and the range of interpretation can thus be extended. The "closed" system, in contrast,

has a fixed number of scales, representing the basic underlying dimensions of the item pool and these will never be expanded unless the item pool is.

The choice between inventories here is a matter of personal choice. Whether one prefers the vast array of the Kuder OIS scales, the stark simplicity of Holland's six scales, or the combination of the SVIB is largely a product of which system one is accustomed to.

As each of these scaling systems were developed differently, the actual scale construction methods are relevant here.

There are a number of ways to build scales--empirical comparisons between criterion groups and Men-in-General, or empirically weighting items according to their popularity within the criterion group, or clustering items with high intercorrelations, or contrasting two criterion groups, or even weighting the items by some intuitive armchair method--how does one decide which method is best? The question of "best" should not be decided by reviewing the scale construction techniques themselves, but rather by reviewing the characteristics of the resulting scales, which is covered in the next point.

SUPPORTING DATA

VII. <u>There should be an extensive body of published information on the inventory to include, at a minimum, extensive normative data, long-range stability data, correlations with other assessment devices, and relationships with outside criteria.</u>

The normative data should include mean scores for a wide range of samples, both students and adults, so that counselors can have some

feeling for what high scores mean. On the pre-1966 Strong, for example, many high school students scored high on the Farmer scale, although they obviously had no intention of choosing farming as a career. When Layton's monograph (1960) showed that over 50 percent of high school boys had high scores on this scale, it became clear that the scale should usually be ignored in counseling young men, because high scores were more indicative of outdoor, adolescent adventuresomeness than of agricultural interests. One can criticize the SVIB--in fact, should criticize it--for having a scale with such a high base rate, but the point here is that the flaw would not have appeared unless this extensive information was available.

Longitudinal studies over long time spans must be carried out to provide interpretive data. From such information, Strong was able to conclude many years ago that interests are very stable after age 25, and from age 15, which is about the earliest that the SVIB can be used, to age 25, the change can be split into thirds, one-third occurring between age 15 and 16, one-third between 16 and 18, and one-third between 18 and 25. Such guidelines are absolutely essential to those working with the inventory.

The basic information for an inventory should be easily accessible, preferably published in one source. No interest inventory currently meets this requirement, though the SVIB probably comes closest. Strong's books (Strong 1931, 1943, 1955), Layton's monographs (Layton, 1958, 1960), Darley and Hagenah's book (1955), the current SVIB Manual (Campbell, 1966) and an extensive Handbook for the SVIB now in press (Campbell, in press)

collectively present an impressive array of background information.

THEORY

VIII. An interest inventory should be tied into a theory which allows users to interpolate beyond the specific test results, that is, to permit interpretation beyond sheer empiricism.

No test or inventory anywhere in psychology achieves this now, as much because psychological theories are inadequate as because the tests are bad. For interest inventories especially, theory, per se, has not had much impact; in the historical context where they were developed, there simply was none.

However, now that the inventories are in existence, a body of knowledge has come into being which in many ways serves a theoretical function. Research has shown that preferences expressed on a paper and pencil inventory are related to actual career choices, that these preferences are stable over time, that they can be scaled, and that the resulting scales can be grouped in meaningful ways. These and other points certainly constitute a substantial body of knowledge that can be used to draw other inferences, and then test them. Although the SVIB was developed empirically, and its author, E. K. Strong, was about as atheoretical a psychologist as could be found, the SVIB now has a large body of supporting data which, when organized according to the rules of philosophy of science, constitutes a moderately respectable theory.

On this point, the Kuder is probably the weakest of the important contemporary interest inventories. For the OIS, there is no guiding theory, no published body of supporting studies, and no attempt to

organize the material into a meaningful framework. For example, on the current OIS profile, the occupations are simply listed alphabetically with no attempt to group them in any manner. Although the Manual says that this is because no groupings emerged clearly from the data, I feel strongly that an investigator has a huge responsibility to organize his data in such a way that some sense emerges, particularly when empiricism can be multiplied beyond reason now by the computer. To simply string out 156 occupational scales on a profile with no underlying rationale is a thinly veiled attempt to out-Strong the Strong--which might be all right except that it is on this particular point that the Strong is most vulnerable.

In using theory to guide both research and development, John Holland's work, with his Vocational Preference Inventory, is superior to anything else being done. His book, the Psychology of Vocational Choice (Holland, 1966) is a thoughtful attempt to integrate what is known about vocational choice into a system that can then be used in test development and for the study of individual development. His theory has not had much impact on psychological research, partially because the current Zeitgeist is not stimulated by research on vocations nor tests, and partially because Holland writes with a distinct lack of jargon. His writing is so clear and literate that psychologists may have ignored it because they felt it had little substance. He has not, for example, used the word "multivariate" even once.

METHODS OF INTERPRETATION

IV. The results of the inventory should be reported in a manner that is easily understandable, that provides the individual with a great

deal of information about himself and how he compares with others, and that leads to the fewest misinterpretations.

This is the most neglected link in interest measurement. We have devoted years and hundred of thousands of dollars to the issues of item format, criterion group composition, scale construction and the like, but the final output that is presented to the counselor and his client has received almost no attention; usually it is dictated largely by the capabilities of the data processing machinery that produces it. As that machinery has become more flexible, the forms are improving, but they are still more determined by practical considerations of printing and mass production than anything else.

On the SVIB profile, attempts have been made to make the scores more useful by grouping the scales into meaningful categories, and by presenting normative data for both the criterion group and Men-in-General. In addition, on the new profile published in 1969, the scores are presented in a way to permit immediate comparison with teenagers, and also with those same men 36 years later as 52 year old adults. This profile provides a fairly decent visual picture of how change takes place, on the average. Still, we have virtually no information on the impact of various forms of output; in general, all interest inventories, indeed all psychological tests, need improvements in the manner that their results are reported.

COMMERCIAL VIABILITY

X. An interest inventory must be commercially viable; that is, it must make money.

There are at least three reasons for this: first, only a commercially successful test can be maintained over a long enough period to collect the necessary supporting data.

In the forthcoming Handbook for the SVIB, test-retest data are presented for samples tested and retested over two weeks, one month, eight months, one year, three years, four years, eight years, ten years, eighteen years, twenty-two years, thirty years, thirty-one years, and thirty-six years. The only possible way to collect such data is to have a commercially viable system which maintains itself, though it is probably also necessary to have a single individual, such as E. K. Strong, who is willing to oversee the system that long.

The second reason for commercial viability is to secure the funds to do the research.

To advance the state of art in any field, one needs money; yet those doing research in the area of psychological testing are faced with problems. The Federal agencies, with some justification, have become schizoid in respect to psychological tests. The National Science Foundation, for example, will not fund any testing projects, seeing them--I guess--outside of the range of science, and the other agencies are running scared as demonstrated, for example, by the Bureau of the Budget's veto power over any test instruments used in projects financed by the National Institute of Health or the Office of Education. As a taxpayer and private citizen, I am in agreement with these policies; I do not have all that much confidence in the propriety of all of my

colleagues--but as a researcher I am very distressed at giving veto power to others who know less about my area than I do.

The private foundations, particularly the Russell Sage Foundation, have supported some work, but the amounts involved are relatively small and more concerned with the impact of testing, and not on testing itself.

Given these problems, about the only dependable source of funds for research is income from the test itself.

The third reason I believe strongly in commercial viability is that this provides some quality checks; no inventory that is invalid, or is in poor taste, or uses some nonsensical approach can maintain itself indefinitely in the market place. The opinions of the profession will affect some change. This type of correction, though it takes some time to operate, is much better than having the Bureau of the Budget looking over your shoulder.

THE PUBLISHER

XI. The "ideal" inventory should be published by an organization that will supply a wide array of ancillary services.

An interest inventory can no longer consist of a booklet, answer sheet, profile, and manual. Interpretive materials should be available, research handbooks containing the basic technical data should be published, case studies should be written up, and an ongoing program of in-service training information should be available for counselors and personnel workers. In addition, the publisher should be involved in the problems of the testing industry, such as test security, invasion of

privacy, unethical actions, and--one of the hot issues of the moment--free exchange of scientific information versus copyright protection. Finally, the publisher should be willing to put up some capital to support research to find new and better ways.

No publisher does all of this well, but the Psychological Corporation does most of them better than anyone else, and they deserve our commendations for their efforts.

Yet there is a quandary here; a prim and proper publisher who attempts to do all of these things precisely according to APA/APGA Standards for Psychological Tests runs the risk of being too cautious and conservative. The Psychological Corporation, for example, is an outstanding publisher, yet the first breakthrough in MMPI scoring, the computer interpreted profile, came not as a result of their efforts, though they have published the test for many years, but from the work of Pearson and his associates at the Mayo Clinic.

From what I know of the activities of test publishers generally, I would say it is highly unlikely that any new psychometric breakthrough will come from any of them. In fact, I am convinced that the next leap ahead in testing will come from an area that most of us highly ethical professionals consider slightly disreputable--the computer dating firms. What they are trying to do is collect systematic data from individuals and then provide these individuals with hitherto unavailable options. The comparable use of psychological test data to expand the individual's horizons has not yet been attempted, but should be.

FINALLY.....

These are, in my opinion, eleven of the most important considerations in evaluating interest inventories. There are others, but they are either specific to a given situation, or related to the above points, or less important.

One feature that inevitably is considered in comparing tests is cost--but this is almost totally irrelevant. The cheapest test available is virtually free; the most expensive one costs about $1.00, so the spread is not great. Any institution that considers $1.00 too much to spend in helping the student think about his career goals is practicing the wrong kind of economy.

The choice between inventories must be based on more substantial considerations; hopefully the above listing will help in these decisions.

REFERENCES

Campbell, D. P. Manual for the Strong Vocational Interest Blanks for men and women. Stanford: Stanford University Press (1966).

Campbell, D. P. HANDBOOK for the Strong Vocational Interest Blank. Stanford: Stanford University Press (in press).

Clark, K. E. Vocational interests of non-professional men. Minneapolis: University of Minnesota Press (1961).

Darley, J. G. and Theda Hagenah. Vocational interest measurement. Minneapolis: University of Minnesota Press (1955).

Holland J. L. The psychology of vocational choice. Waltham, Mass.: Blaisdell Publishing Company (1966).

Layton, W. L. Counseling use of the Strong Vocational Interest Blank. Minneapolis: University of Minnesota Press (1958).

Layton, W. L. The Strong Vocational Interest Blank: research and uses. Minneapolis: University of Minnesota Press (1960).

Perry, D. K. Forced-choice vs. L-I-D response items in vocational interest measurement. Journal of Applied Psychology, 1955, 39, 4, 256-262.

Strong, E. K. Jr. Change of interests with age. Stanford: Stanford University Press (1931).

Strong, E. K. Jr. Vocational interests of men and women. Stanford: Stanford University Press (1943).

Strong, E. K. Jr. Vocational interests 18 years after college. Minneapolis: University of Minnesota Press (1955).

Zuckerman, J. V. Interest item response arrangement as it affects discrimination between professional groups. Journal of Applied Psychology, 1952, 36, 2, 79-85.